

# The Hirsch-index: a simple, new tool for the assessment of scientific output of individual scientists

## *The case of Dutch professors in clinical cardiology*

T. Opthof, A.A.M. Wilde

In this brief paper we explore the Hirsch-index together with a couple of other bibliometric parameters for the assessment of the scientific output of 29 Dutch professors in clinical cardiology. It appears that even within such a homogeneous group there is large interindividual variability. Although the differences are quite remarkable, it remains undetermined what they mean; at least it is premature to interpret them as differences in scientific quality. It goes without saying that even more prudence is required when different fields of medicine and life sciences are compared (for example within University Medical Centres). Recent efforts to produce an amalgam of scientific 'productivity', 'relevance' and 'viability' as a surrogate parameter for the assessment of scientific quality, as for example performed in the AMC in Amsterdam, should be discouraged in the absence of a firm scientific base. Unfortunately for politicians and 'managers of science' only reading papers and studying are suitable for quality assessment of scientific output. Citations analyses can't substitute that. (*Neth Heart J* 2009;17:145-54.)

Keywords: citation analysis, clinical cardiology, grant proposals, Hirsch-index, impact factor, peer review

### T. Opthof

Department of Experimental Cardiology, Center for Heart Failure Research, Academic Medical Center, Amsterdam and  
Department of Medical Physiology, University Medical Center Utrecht, Utrecht, the Netherlands

### A.A.M. Wilde

Department of Cardiology, Academic Medical Center, Amsterdam, the Netherlands

Correspondence to: T. Opthof  
Department of Experimental Cardiology, Center for Heart Failure Research, Academic Medical Center, PO Box 22660,  
1100 DD Amsterdam, the Netherlands  
E-mail: t.opthof@inter.nl.net

The format of scientific papers is essentially unchanged since 1850, despite formidable changes in science itself. These changes relate to the absolute number of scientists, to technology and to the turnover of new information. The old format of scientific papers has even survived the era of electronic publishing. There are only minor differences in format between disciplines within the life sciences, including medicine. In medical papers the Discussion section is somewhat long compared with papers in, for example, physics or mathematics. Also, there are differences in the total number of references in a scientific paper, which have an impact on 'the game' of citation analysis.

### Citation in science

Two versions of the same scientific paper, one with and one without references, bear the same amount of information. In the version with references we offer the readership and the peers the possibility to check information. Of course, this is only a theoretical issue, because scientists do not have time and possibly also not the interest to do this in practice. Reading a whole paper in detail rather than browsing the abstract is already a time-consuming task. Even reviewers of submitted manuscripts can in practice only afford to check a couple of references at most. Scientific writing therefore remains a matter of trust. Pressure of 'managers of science' to publish more and more and in journals with an impact factor as high as possible, is a potential threat to this basic trust as we see now and then in remarkable cases of fraud and misconduct. Whether or not the recent Woo-Suk Hwang debacle constitutes the tip of an iceberg is unknown.<sup>1</sup> Checking by authors how their own work is referred to by colleagues is sometimes a frightening experience and its digestion requires a strong stomach.

When the results of a study are questionable or whistling against the wind, the role of references becomes even more important. In general, there is an urge for editors to select papers for publication with a

high preference for novelty. Any experienced author who has compiled a review paper on whatever topic, will acknowledge the fact that it is in retrospect completely unimportant which group of authors published certain data first. In fact, it is more worthwhile that data are published by different independent groups – in whatever temporal order – than that they are published by one group and confirmed by the same group again and again with incremental information from paper to paper.

The order by which different groups publish conflicting data is highly relevant to an important characteristic of science. Individual scientists may not appreciate reading that science is conservative. First published data do have a tailwind that makes it less easy for subsequent authors to publish conflicting material. Although this may seem contra-productive in terms of progress, it is useful that existing knowledge is protected to a certain degree by this conservatism. The onus for providing scientific overweight to new insights over existing ones is on future authors. This can be quite painstaking, as Galilei found out.

The emphasis on novelty rather than confirmation or rebuttal, in particular by editors of major journals, may provoke authors to 'polish' their data. This may explain why frequently cited (clinical) research is not always robust enough to withstand future challenges. Ioannidis<sup>2</sup> recently analysed 49 original clinical research studies that were published between 1990 and 2003 and that were cited over 1000 times each until August 2004. Of these, 45 had been claiming an effective intervention. Given the fact that it concerned studies with large numbers of groups and scientists, it is quite remarkable that the outcome could not be confirmed in 31% of these studies with half of the conflicting analyses reporting smaller effects and the other half reporting even contradictory results.<sup>2</sup> In fact, only a minority (44%) was fully confirmed.

Studies with a positive outcome can more easily be published than studies with a negative outcome, in particular in trials. This publication bias is a well-recognised problem and a potential danger to meta-analyses in which probabilities derived from trials or efficacy of pharmaceutical agents are summarised and weighed.

Citing previous work has an important role in providing a solid base to whatever scientific study. In addition, it has a social role. Any discipline has founding fathers and contemporaneous leaders. By citing their work, scientists make themselves acceptable for their peers who may also be reviewers of their manuscripts. It is not possible to publish new material whatever its quality without demonstrating a minimal overlap with the status quo by including relevant references to reach this aim. Subsequently, the study may deviate from existing views. This conservatism in science may be stronger than in the arts, but this seems quite acceptable when public health is involved. Given the fact that citation also has a social role, it follows that

frequent citation gives prestige to the cited author, including his establishment.

### Does citation reflect scientific quality?

In the past, an attempt was made to correlate 'peer esteem' with the 'citation rate' of individual authors.<sup>3</sup> Questionnaires provided the 'peer esteem'. The fields of biochemistry, psychology, chemistry, physics and sociology were analysed and the correlation coefficients varied between only 0.53 and 0.70.<sup>3</sup> This implies that only 25 to 50% of the differences in peer judgment about scientific groups were associated with differences in citation scores in that study.<sup>3</sup>

The answer to the question heading this section is positive for top scientists. Cole<sup>3</sup> showed in 1989 that Nobel Prize winners are more frequently cited than other scientists and this difference is not caused and not even quantitatively affected by the award. Nobel Prize winners constitute a special 'hors categorie', to lend a phrase from the Tour de France. It has not been established whether differences in citation rate among scientists in general is a suitable tool for defining differences in scientific quality.<sup>4</sup> In the absence of such a relation, one should in fact refrain from ranking scientists on the basis of citation data. The human interest to know one's position in a competitive surrounding has given ground to citation analysis. In the next section we will see that there are additional parties interested in citation analysis.

### Citation analysis

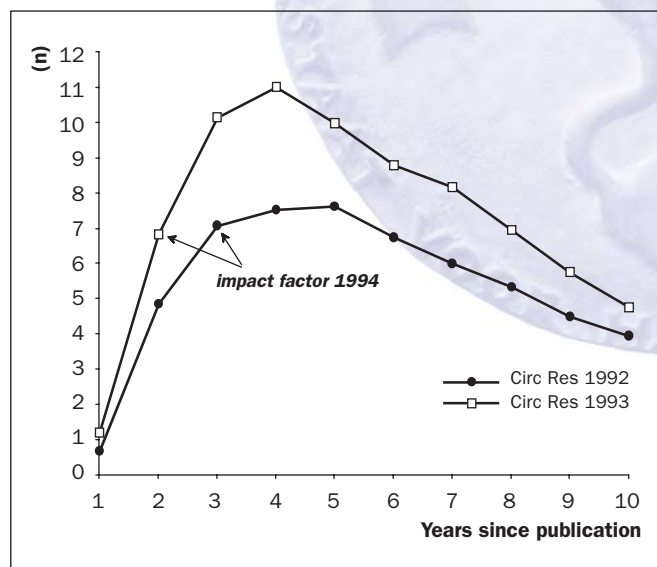
Assessment of scientific output by citation analysis has raised the interest of specialised scientists and has evolved into an independent scientific discipline with its own specific journals. Citation analysis serves a number of goals. First, it may uncover the way by which scientific information spreads through the community. This in itself has already been attracting attention from the social sciences since the second half of the previous century. Second, bibliometric parameters may help librarians to select scientific journals. This has become imperative because of budget constraints. These are relevant for the majority of academic institutes, because institutional subscriptions for scientific journals are expensive, despite the fact that both editing and reviewing of scientific journals are largely paid by the academia ('scientific volunteers'). The fact that such professional consultancy is free is quite remarkable and it is a great asset. A third goal is in the ranking process itself. Research increases knowledge in general, and medical research has an additional role in improving public health as a long-term goal, but the work itself is done by humans. Therefore, the scientific arena is subject to human behaviour, including extremes as we have recently seen in the cloning debacle around Woo Suk Hwang in South-Korea.<sup>1</sup> However, competitive behaviour in itself is a normal biological characteristic and thus competitive listings of scientific output satisfies biological needs. Fourth, financial resources are not

endless. For this reason, institutions involved in assessment of grant proposals such as the Netherlands Organisation for Scientific Research (NWO) and the Netherlands Heart Foundation (NHS) are interested in numerical parameters that can add credibility to their decision making system, primarily based on peer-review.

### Impact factors

The impact factor is a parameter relevant for scientific journals, not for individual papers and certainly not for individual scientists. The impact factor of a scientific journal is the average number of citations obtained in a given year by the papers published in the two years preceding the year during which these citations were obtained. The *Netherlands Heart Journal* is about to obtain an impact factor for the first time in 2009. It will be calculated by summation first of the total number of citations obtained in all scientific journals during 2009 to all papers published in 2007 and 2008 by the *Netherlands Heart Journal* and dividing this sum by the total number of papers published.

Citation of papers has an impact on the position of both individual scientists and on the journals in which they publish their work.<sup>4,5</sup> Figure 1 shows citation of the contents of *Circulation Research* as published in 1992 and 1993. Along the abscissa years since publication are depicted ranging from 1 to 10. For the contents of 1992, year 1 is 1992 and year 10 is 2001. In a similar way for the contents of 1993, year 1 is 1993 and year 10 is 2002. Along the ordinate we find the average citation of the complete contents of *Circulation Research* in 1992 and 1993, respectively.



**Figure 1.** Citation of papers published by *Circulation Research* in 1992 and 1993. Abscissa: years 1 to 10 are 1992 to 2001 for the 1992 contents and 1993 to 2002 for the 1993 contents. The impact factor of 1994 for *Circulation Research* is the weighed average of year 2 for the 1993 contents and year 3 of the 1992 contents. Ordinate: citation per paper during the ten years since publication.

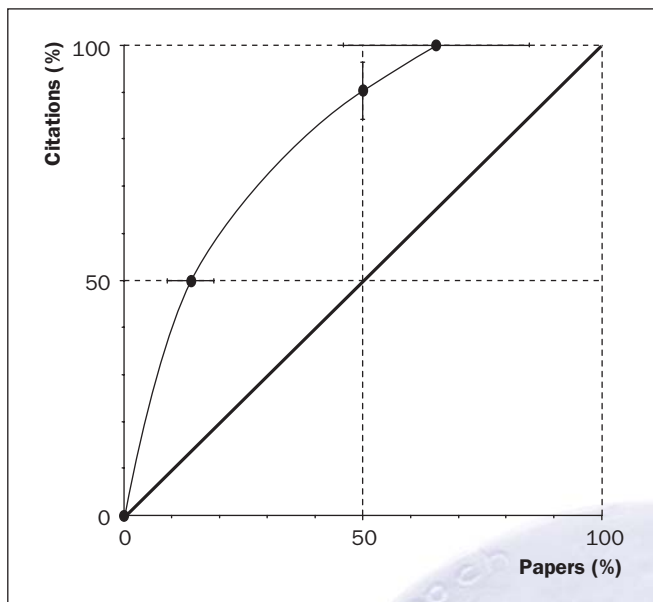
The contents of 1992 contributed to the impact factor of *Circulation Research* for 1993 and 1994. The contents of 1993 contributed to the impact factor of *Circulation Research* for 1994 and 1995. The impact factor of *Circulation Research* for 1994 is the weighed average of citation of the contents of 1992 as cited in year 3 and citation of the contents of 1993 as cited in year 2, as indicated by the two arrows in figure 1.

Apart from explaining calculation of the impact factor, figure 1 also shows the time period over which citation of the large majority of papers normally increases, peaks and subsequently wanes. The period during which papers are most frequently cited lags behind the period relevant for the calculation of the impact factor. In the cardiovascular sciences this 'zenith' is often in year 4 or even in year 5. Figure 1 shows that in year 9 or 10, citation may still be as frequent as in year 2. In more recent years there is a trend to a shift of the zenith to year 3, in particular in clinically oriented journals such as *Circulation* (not shown). Whether this is caused by temporary trendy research topics or by the availability of electronic publishing with its concomitant retrieval tools, is unknown. The average citation curve in figure 1 does not apply to all papers. There are papers that follow a completely different time course as has been described previously.<sup>6</sup> Such papers are in a certain sense 'timeless'. They escape from the citation profile as described in figure 1. In our opinion this is an important parameter of scientific quality. Science managers, however, 'do not have time' to appreciate this.

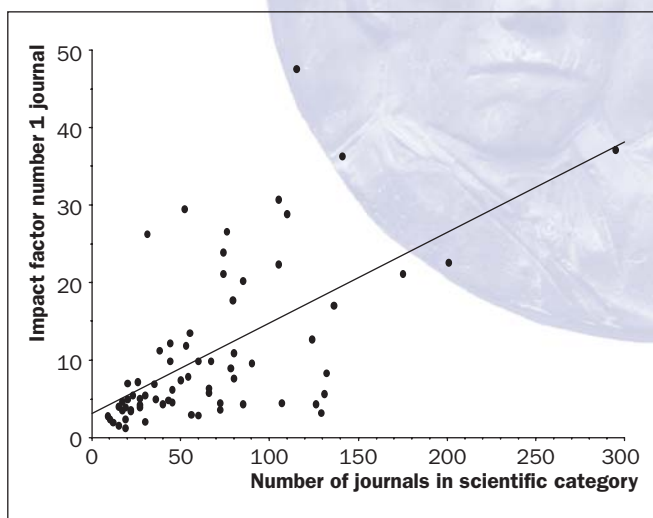
### Limitations of the impact factor

Often the impact factor of a journal in which a paper has been published is used as some kind of parameter of esteem for that paper, for example by committees that compare the scientific output of individuals or departments. This practice has been amply criticised by numerous scientists,<sup>3,4,7-9</sup> but still continues. Figure 2 explains why an individual paper cannot serve as a *pars pro toto*.<sup>10</sup> The abscissa shows a survey of a large number of datasets from *Cardiovascular Research* collected between 1992 and 2002 and cited during subsequent years. In each set of data the papers were ranked along the abscissa from most frequently cited to less frequently or not cited and transformed to a scale from 0 to 100%. The same was done along the ordinate for the number of citations obtained by those papers. The end of the dashed line of identity in the upper right corner depicts the total number of citations obtained by all papers. Only this point in the graph is comparable to the impact factor and it would be visible as such if the scales of both axes were not in percentages, but in real numbers, simply by dividing the value at the ordinate by the value at the abscissa for the relevant years of publication and citation. Figure 2 indicates, however, that citation is heavily skewed. Thus, in this case, only 14% of all papers obtain 50% of all citations. Also, the





**Figure 2. Skewness of citation.** When papers are ranked according to citation order along the abscissa and plotted against the number of citations along the ordinate (both in percentages), the skewness of citation becomes obvious. It concerns 63 sets of original papers ( $n=1886$ ) as published between 1992 and 2000 by Cardiovascular Research and analysed for citations between 1992 and 2002. About 14% of papers obtain 50% of citations. The most frequently cited 50% of papers obtain about 90% of all citations. Only 66% of papers are cited in a given year, leaving 34% uncited. Obviously, the impact factor of a journal cannot be regarded as a 'totum pro parte' for individual papers, although this is done by many academic institutions. Compiled from reference 10.



**Figure 3. The impact factor of a number 1 ranked journal within 69 (sub)disciplines** (such as 'Cardiac & Cardiovascular Systems' or 'Biochemistry & Molecular Biology') in life and medical sciences is correlated with the total number of journals within that category. Data are taken from ISI Journal Citation Reports® 1999 (Thomson Corporation). There are large differences between (sub)disciplines. Compiled from reference 15.

most frequently cited half of the papers obtain about 90% of all citations. A considerable number of papers remain uncited for many of the years after publication. Details of this analysis have been described previously.<sup>10</sup> Assignment of the impact factor of a whole journal to an individual paper will overestimate the actual number of citations of 80 to 90% of papers and it will seriously underestimate the number of citations of the remaining 10 to 20% of papers. In summary, it describes the obtained citations of journals as a whole, but of *none* of the individual papers. With this in mind, it is remarkable, to put it mildly, that the parameter is still used for quality assessment of scientific output.<sup>4,8,9,11</sup>

Apart from the skewness of citation, the number of scientists involved, exemplified by the number of scientific journals within a scientific category, determines the impact factor of the top journals and therefore also of the papers belonging to that category. Figure 3 shows the impact factors of 69 journals from the medical and life sciences as published in the 'ISI Journal Citation Reports®' of 1999. ISI Journal Citation Reports® is one of the products of the Thomson Corporation, formerly known as the Institute for Scientific Information. These 69 selected journals all concern the top journal of a specific discipline, such as 'Cardiac and Cardiovascular Systems'. Obviously, the number of journals within a discipline is correlated with the impact factor of the leading journal in that same discipline. The logic behind this is that there are simply more citations to obtain within a (sub)discipline with a lot of active scientists. The consequence is that science watchers in academic institutions should not compare groups within their institution with each other, but with comparable groups in other institutions at an international level.<sup>11</sup> In some institutions this hurdle is circumvented by regarding a fixed percentage of journals within a category as top journals. When such a percentage is, for example, set at 10%, papers in the top 6 of a category with 60 journals would be regarded as top papers, whereas this qualification would require publication in the top 25 of a category with 250 journals. Whatever the approach, problems as indicated in figure 2 on skewness of citation, remain.

### The *h*-index: an improvement?

Recently, Hirsch proposed the *h*-index as an alternative parameter for measuring the scientific output of an individual.<sup>12</sup> Hirsch stated 'I propose the index *h*, defined as the number of papers with citation number higher or equal to *h*, as a useful index to characterise the scientific output of a researcher' (see reference 12, but also [www.arXiv.physics/0508025](http://www.arXiv.physics/0508025)). As often with good ideas, the method is surprisingly simple. Figure 4 shows the calculation of the *h*-index. All cited papers of an author are ranked along the abscissa in the order of the frequency by which they have been cited. The vertical and horizontal lines at '*h*' along both axes cross at '*h*,*h*'. Let us assume that *h* amounts to 3 in this case.

The horizontal line indicates the theoretical situation that this author has written many papers (e.g. 100) that all have been cited three times. The vertical line indicates another hypothetical situation, i.e. the author has published three papers each of which may have been cited very frequently (e.g. 100 times each). In both cases the *h-index* is 3. The method is attractive because it appreciates both productivity and frequent citation. Moreover, it mitigates the effect of outliers, e.g. produced by a large clinical trial in which an author may be in the midst of a large string of authors without a major intellectual contribution to the paper. Also, papers that remain uncited, e.g. because they were published with an educational aim rather than a scientific aim, do not burden the *h-index* of an individual. There are many more practical advantages. One does not need to know a full publication list of an author, because only cited papers are of importance. Also, one only needs entrance to a database such as the ISI Web of Science® without dependence from additional, derived bibliometrical products as ISI Journal Citation Reports® which not only saves money but also time because impact factors are published with considerable delay. At a more local level, one does not need committees, including supporting staff, that try to invent wheels that may not exist.

Figure 4 shows in its inset also a formula ( $C=ah^2$ ; it is probably not a coincidence that this novel method was developed by a physicist...), which estimates the total number of citations of an author from his or her *h-index*. In this formula 'C' is the total number of citations and 'a' is a constant which varies between 3 and 5 as has been established empirically by Hirsch.<sup>12</sup> Figure 5 shows the track record of two scientists each with a *h-index* of 3. Author 1 has two papers which were both cited ten times and an additional eight papers with 3 citations each. The total number of citations is 44 and the *h-index* is 3, because the 3rd paper was cited 3 times, but the 4th paper was *not* cited 4 times. The second author has three papers cited 5, 4 and 3 times and an additional seven papers which were cited once each. This author with a total number of citations of 19 is also assigned an *h-index* of 3. The relevant area in the graph for the *h-index* of both authors has been filled in in figure 5. During a career the *h-index* will increase. It cannot decrease. The temporal aspects of its development remain to be established. Although thus far it has not been mentioned by specialists in this field, or by Hirsch himself, it seems useful to make a distinction between the *h-index* of the total output of a scientist and his or her first authored papers. We will perform such an analysis on all Dutch professors in clinical cardiology later in this paper. The remarkable simplicity of the proposal by Hirsch<sup>12</sup> has, as could be expected, provoked a plethora of subsequent work with refinements. For those interested in more details we refer to a paper by Bornmann et al.,<sup>13</sup> in which nine variants of the *h-index* are compared. A perfect example of making simple things intricate....

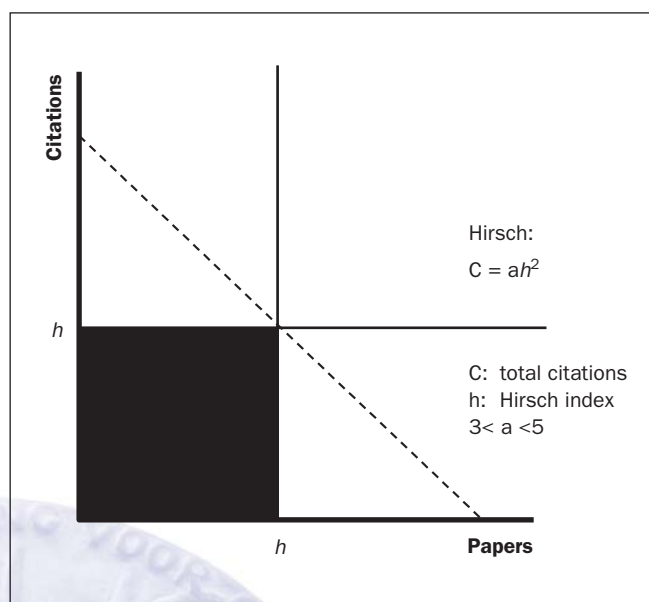


Figure 4. Explanation of the Hirsch (*h*)-index. 'C': total number of citations; 'a': constant which is between 3 and 5 according to empirical analysis. 'h': number of papers with citation number higher or equal to *h*. See text, but also reference 12 for further explanation.

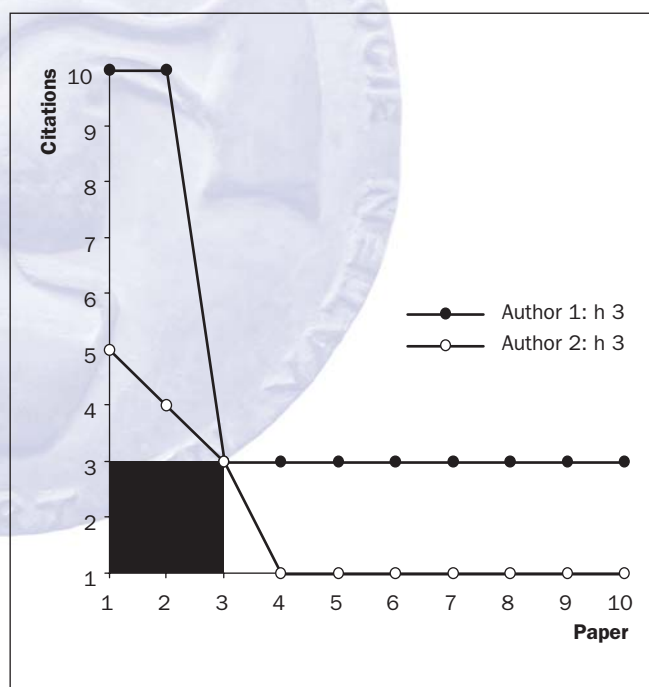


Figure 5. Example of *h-index* of two (non-existing) authors. Both authors have written ten cited papers. One has a total of 44 citations (author 1), the other a total of 19 citations. Both obtain *h-index* 3, because the 3rd paper is cited three times and the 4th paper is not cited four times (see the filled area). Note that following the formula  $C = a \cdot h^2$  (see figure 4), the expected total citations would be between 27 and 45. The profiles of these two ghost authors are therefore extreme examples.

A serious limitation of the *h-index* is that it favours those with a long career. This is not attractive for scientists at the beginning of their career. There are two practical solutions for this 'problem'. One is to limit the citation/publication period to a fixed number of recent years. For example, a five-year period, which is updated year by year: 2004-2008, 2005-2009, 2006-2010 and so on, although this assumes that the accumulative value of the *h-index* is more or less linear. Another alternative is to limit the assessment to first authored papers. This will also bring junior scientists in a more favourable position compared with senior scientists. It also makes clear whether or not senior scientists remain active themselves other than by steering younger scientists or providing them with facilities in exchange for authorships. For senior (last) authorship it is not possible to distinguish between authorship with a relevant intellectual contribution or honorary 'ghost authorship'. Another major advantage is that in such a system each paper is only counted once. Recent developments of assessment of the scientific output of so-called principal investigators (PIs), as proposed in the Academic Medical Center in Amsterdam,<sup>11</sup> completely blurs the output of scientific groups as a whole, let alone of an academic institution, because the overlap between the publication lists of the individuals may be considerable and invisible as well ('strategic publishing').

### The *h-index* and individual scientists

Table 1 shows a listing of the *h-indices* of prominent physicists (left) and of top scientists in the life sciences (right). It is an assessment of all papers authored by these individuals. Ed Witten from the Princeton Institute for Advanced Study, who devised the M-theory, is at rank 1 amongst the physicists. Nobel prize winners Philip Anderson and Pierre-Gilles de Gennes take positions 5 and 9. The numbers for the physicists were taken from a brief paper of Ball, dating back to 2005.<sup>14</sup> The numbers for the prominents in the field of life sciences are derived from the paper by Hirsch.<sup>12</sup> In comparing *h-indices* of individuals it should be taken into account that it is in fact a squared parameter. Thus, the difference between 19 and 20 is as large as that between 5 and 8. By comparing the left and right part of table 1 it is obvious that *h-indices* are about 50% higher in the life sciences than in physics. In figure 3 we have shown that even within a field like medicine and life sciences the impact factor varies with the number of journals (and probably with the number of active scientists; see also reference 15). Thus, the impact factor of the leading journal in the category Biochemistry & Molecular Biology is substantially higher ( $\approx 31$ ) than that in the category Cardiac & Cardiovascular System ( $\approx 12$ ). Without doubt this also plays a role when one compares *h-indices* of different subfields within a category. It can be anticipated that papers on, for example, computer modelling of cardiac electrical activity will acquire less citations than papers

**Table 1.** The *h-index* of top scientists in physics and life sciences.

Prominent physicists		Life sciences	
Name	<i>h-index</i>	Name	<i>h-index</i>
E. Witten	110	S.H. Snyder	191
A.J. Heger	107	D. Baltimore	106
M.L. Cohen	94	R.C. Gallo	154
A.C. Gossard	94	P. Chambon	153
P.W. Anderson	91	B. Vogelstein	151
S. Weinberg	88	S. Moncada	143
M.E. Fisher	88	C.A. Dinarello	138
M. Cardona	86	T. Kishimoto	134
P.G. de Gennes	79	R. Evans	127
S.W. Hawking	62	P.W. Serruys	74/95

Note that the rating of prominent scientists from the life sciences is higher than that of physicists. Data were taken from references 12 and 14 and were determined in 2005. We added P.W. Serruys's score; the figure before the slash (74) is pertinent to 2005, whereas the figure after the slash (95) is per 1 June 2008. See also table 2.

on heart failure, ischaemia, myocardial infarction, or atrial fibrillation, to mention just a couple of larger fields of research, simply by the difference in scientists involved in the fields. Within the life sciences (right part of table 1) the top position is taken by S.H. Snyder with *h-index* 191. For comparison we have added the prominent Dutch clinical cardiologist P.W. Serruys with *h-index* 74 in 2005 to this list. The number after the slash shows the *h-index* by 1 June 2008 (95) (see also table 2).

As pointed out by Hirsch<sup>12</sup> the *h-index* increases with time and it cannot decline. Hirsch predicted that it increases linearly with time, but this is based on mathematics and not on data (analysis of careers). It would be of interest to monitor the *h-indices* of individuals during the development of their careers, for example, for present post-docs. The increase of the *h-index* of P.W. Serruys from 74 in 2005 to 95 in 2008 suggests that its development is not linear in the field of clinical cardiology. It has been proposed to use this type of numerical data for appointments with an important academic prestige. One might think of 'tenure', 'established investigators', 'professors' and (in the Netherlands) 'members of the Royal Netherlands Academy of Arts and Sciences (KNAW)'. The fact that the *h-index* can only increase and not decrease does, of course, limit its significance for assessment of current performance. It is, however, a very suitable parameter for the assessment of a long career. In the previous section we explained that there are ways to adapt the *h-index* in such a way that it can also monitor current performance.



**Table 2.** Hirsch (H)-indices of 29 Dutch professors in clinical cardiology (1 June 2008).

Author	Affiliation	h-index All papers	Rank	h-index 1st author papers	Rank	Papers (n)	Citations (n)	Rank	Cit./paper (n)	Rank	m (h/yr)	Rank	First year	A (C/H2)	Comment
Arnold AE	MCA	18	27	6	24	37	2351	21	63.54	1	0.86	24	1988	7.26	Edited
Bax JJ	LUMC	37	9	18	4	620	7431	7	11.99	27	2.85	2	1996	5.43	Edited
Crijns HJGM	AZM	43	6	12	9	347	8274	6	23.84	15	1.48	13	1980	4.47	Edited
De Feyter PJ	Erasmus	63	3	20	2	428	15346	3	35.86	5	1.97	5	1977	3.87	
De Winter RJ	AMC	19	26	12	9	95	1286	27	13.54	26	1.27	18	1994	3.56	
Doevendans PAFM	UMCU	30	13	7	22	151	2818	18	18.66	18	1.88	7	1993	3.13	
Gorgels APM	AZM	23	23	9	17	87	1619	26	18.61	19	0.74	28	1978	3.06	
Hauer RNW	UMCU	29	14	8	20	92	3180	15	34.57	6	1.07	21	1982	3.78	Edited
Jordaens LJ	Erasmus	22	24	10	14	176	3601	14	20.46	16	0.79	27	1981	7.44	
Jukema JW	LUMC	34	11	9	17	206	5198	12	25.23	12	2.00	3	1992	4.50	Edited
Mulder BJM	AMC	17	28	2	28	111	1103	28	9.94	28	0.81	26	1988	3.82	Edited
Peters RJG	AMC	21	25	6	24	87	2259	22	25.97	11	0.88	23	1985	5.12	Edited
Piek JJ	AMC	27	17	8	20	152	3650	13	24.01	13	1.50	11	1991	5.01	Edited
Pijls NHJ	Catharina	29	14	11	12	92	2656	19	28.87	9	1.32	16	1987	3.16	
Pinto YM	AMC	24	21	9	17	107	1932	24	18.06	22	1.26	19	1990	3.35	
Schalij MJ	LUMC	27	17	4	26	208	2819	17	13.55	25	1.29	17	1988	3.87	
Serruys PW	Erasmus	95	1	36	1	902	38618	1	42.81	4	2.97	1	1977	4.28	
Simoons ML	Erasmus	68	2	20	2	463	24974	2	53.94	2	2.00	3	1975	5.40	
Smeets JLRM	Radboud	26	19	3	27	84	2528	20	30.10	8	0.84	25	1978	3.74	
Van der Giessen WJ	Erasmus	38	8	10	14	182	5684	9	31.23	7	1.41	14	1982	3.94	
Van der Wall EE	LUMC	48	4	11	12	610	8418	5	13.80	24	1.60	9	1979	3.65	Edited
Van Rossum AC	VU	26	19	8	20	113	2177	23	19.27	17	1.18	20	1987	3.22	Edited
Van Veldhuisen DJ	UMCG	37	9	11	12	383	7033	8	18.36	21	1.95	6	1990	5.14	
Verheugt FWA	Radboud	48	4	14	7	275	13082	4	47.57	3	1.50	11	1977	5.68	
Visser FC	VU	27	17	7	22	168	3110	16	18.51	20	0.96	22	1981	4.27	Edited
Von Birgelen C	MST	24	21	15	5	119	1886	25	15.85	23	1.60	9	1994	3.27	
Waltenberger J	AZM													NA	
Wilde AAM	AMC	43	6	14	7	202	5588	10	27.66	10	1.72	8	1984	3.02	
Zijlstra F	UMCG	31	12	14	7	219	5230	11	23.88	14	1.35	15	1986	5.44	Edited
<b>Average</b>		<b>39.3</b>		<b>13.2</b>											

The *h*-indices of all papers and of 1st authored papers as of 1 June 2008. Ranking follows each column in italics. Individuals with more than one initial were also scored for initials less than their total. On average, these Dutch professors score 39.3 on all papers and 13.2 on 1st authored papers. This analysis is complete for the years since 1975 and is based on the ISI Web of Science® in General Search mode. For the years prior to 1975 the Cited Reference Search mode was added. However, this only considers first authored papers. Therefore the 'All' score of individuals with an important 'oeuvre' before 1975 may have been underestimated by this analysis. We were in doubt about preceding affiliations of one of the individuals. Clarification of this uncertainty required personal contact and because we have chosen not to do this in order to keep the procedure simple, we have not included these data for one individual. For other parameters we refer to the text. Parameter 'A': see figure 4. 'Edited' means that the list of publications of an individual had to be edited by hand, because there were more individuals with the same name. NA=not analysed.

### The *h-index* and Dutch professors in clinical cardiology (past performance)

Table 2 shows a listing of the 29 Dutch professors in clinical cardiology appointed at the eight academic hospitals and an additional three peripheral hospitals. The *h-index* of their total scientific output and of their output based on first authorship is indicated together with a ranking in italics in the columns directly at the right of each listed parameter. These numbers are pertinent as to 1 June 2008. They cannot, therefore, be easily retrieved/checked from the source (Web of Science). On average, the *h-index* is 39.3 for the total output and 13.2 for the first authorships (see for technical details on the calculation the legend of table 2). There are large differences, but the *h-index* does not take into account the years of scientific activity. Even if an effort were to be undertaken to correct for this,<sup>16,17</sup> it would not be obvious which year should be taken as starting point. This could be the year of publication of any paper, cited or not, or of a first authored paper or even appointment at an institution, or the year of publication of a thesis. Because we refrain from interpretation of differences in terms of scientific quality at this stage, the listing is alphabetical. Moreover, it should be realised that the *h-indices* of these individuals may partly be based on identical papers. It is, therefore, not possible to construct the total performance of, for example, the five professors from Erasmus-Rotterdam from their individual scores, because it would require extensive editing of the lists of papers relevant for the *h-index* of these individuals. The *h-index* is more suitable for individual assessment, rather than group assessment, unless only first authorships are considered. Van Raan<sup>18</sup> has stated that groups are the ideal level of quality assessment, but this requires a definition of what is actually a group. At least one would like to correct for the number of scientists involved, which is completely neglected in such studies.<sup>18</sup> Different levels of overlap between the papers of individuals in the eight affiliations (there is considerable overlap at Erasmus-Rotterdam and LUMC-Leiden, but less overlap at AMC-Amsterdam) suggests that group assessment will show smaller differences than the assessment between individuals as listed for the *h-index* based on all papers in table 2. Of course, comparisons other than by the *h-index* are possible. For this reason we have also listed the total number of published papers, the total number of citations and the averaged number of citations per paper. We have only included original papers, reviews and editorials and excluded meeting abstracts and letters. When citations per paper are considered as parameter it is obvious that there are several individuals with much lower ranking, caused by relatively many published, but hardly cited papers. This may point to local differences in publication behaviour, which may either be traditional or in response to pressure from the management. Table 2 also divides the *h-index* by the time between the publication year of the first cited paper and 2008. Hirsch<sup>12</sup> has previously pointed to

values between 2 and 3 as a parameter for excellence. For what that is worth, some individuals have *h-indices*/year above 2.

The parameter 'A' (see inset in figure 4) was estimated by Hirsch to be between 3 and 5 in general.<sup>12</sup> The larger 'A' is, simply means that the *h-index* progressively underestimates the total number of citations. In this subgroup of Dutch professors in clinical cardiology 'A' is never less than 3, but 9 out of 28 times higher than 5.

Table 2 shows that the differences in *h-indices* between individuals are substantial both based on all and on first authored papers. This is of interest for two reasons. First, it gives a clue to young clinical scientists of what seems to be a minimal requirement for reaching the rank of professor in clinical cardiology as far as the scientific part of the job is concerned. Second, the past performance of a scientist in clinical cardiology as a first author appears a rather accurate estimate for his/her total oeuvre and thereby a predictor of future performance. Thus, figure 6 shows the relationship between the *h-index* based on the first authored papers and the *h-index* based on all papers for the same individuals as in table 2. With a correlation coefficient of 0.886 ( $n=28$ ,  $p<0.0005$ ), its square indicates that the past performance as a first author predicts for over 75% of the variability in the *h-index* based on all papers between the individuals in table 2. The remainder of this variability is probably determined by factors as strategy (creating networks) and differences or changes in field of research. For some individuals there are remarkable and rather sudden changes in citation frequency of their work (not shown). For 'managers of science' this implies that a strong position as first author does predict a high *h-index* in the future.

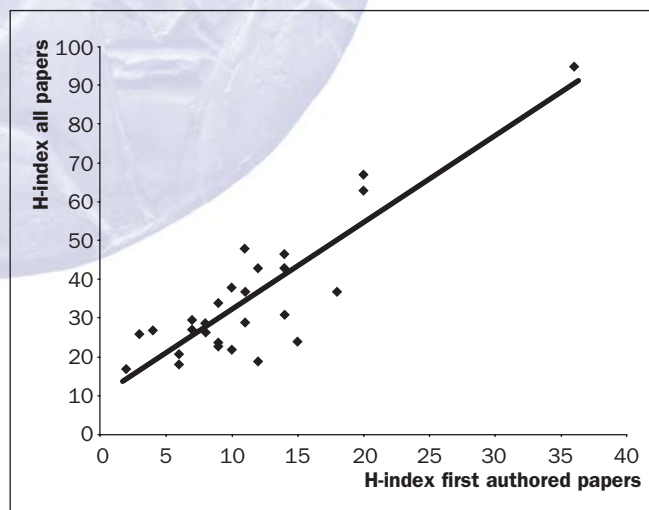


Figure 6. The *h-index* of 28 Dutch professors in clinical cardiology based on first authored papers vs. *h-index* based on all papers. Numerical data can be found in table 2. Citations were counted from data of publication until 1 July 2008.  $Y=2.25 X + 9.6$ ;  $r=0.886$ ;  $n=28$ ,  $p<0.0005$ .



### Peer review and citation data

Reviewers of accepted manuscripts are not capable of predicting future citation in a relevant way. An analysis performed by *Cardiovascular Research* showed that priority ratings of individual reviewers determine less than 2% of the variability in citation of papers during the three years directly following publication of the papers (analysis based on 36 months, i.e. not based on calendar years).<sup>19</sup> Therefore it does not come as a surprise that peer-review assessment of the quality of research groups and of grant proposals also does not match very well with citation counts with whatever method obtained, as has been demonstrated at an international scale,<sup>3</sup> but also in the Netherlands in the past<sup>20</sup> and more recently.<sup>21,22</sup> It does not appear possible to match peer reviewer's qualifications as 'very good' or 'excellent' with significant differences in citation parameters. This constitutes a major limitation for the allocation of research grants by institutions as NWO<sup>23</sup> or the Netherlands Heart Foundation, but also for local scientific quality assessments (see below).<sup>11</sup>

Van Raan compared a so-called crown-parameter with the *h-index*.<sup>22</sup> The study was primarily performed in order to test the robustness of the crown-parameter developed by the Center for Science and Technology Studies (CWTS, Leiden, the Netherlands) against the *h-index*. However, the most important finding of that study is, although not appreciated by the author, that in a comparison of 12 chemistry groups of the same university the qualifications 'very good' and 'excellent' of reviewer panels scored equally high, regardless whether the crown-parameter or the *h-index* was used. Thus, 'very good' and 'excellent' cannot be differentiated from each other based on citation data.

### Selection of grant proposals and citation data

The inference of the lack of correlation between the peer-review rating of grant proposals and citation data is that decision making on such proposals cannot be substantiated with relevant citation parameters afterwards. The steering of such processes, for example by the choice of reviewers, turns the procedure more or less into a lottery for the applicants.<sup>24</sup> An important analysis by Van den Besselaar and Leydesdorff<sup>25</sup> of Dutch NWO programmes including 'Veni', 'Vidi' and 'Vici' has unambiguously shown that there are no differences between granted and the best non-granted research proposals as far as numerical citation parameters are involved. The peer-review system is solid for recognising work that should not be granted, but is poor, if not unsuitable, for making a distinction between what is very good or excellent. This renders the whole process subject to personal bias. Therefore, it is important for scientists to be a member of committees which decide on the allocation of research budgets. Despite the quality rating systems used by peer reviewers, and averaged and discussed by research boards, final decision making is ultimately based on politics, not on objective citation data. This is not

saying that the process is corrupt. It is very much the same as with democracy. The system has pitfalls and limitations, but it is difficult to find better alternatives.

Although there is no alternative for peer review of grant proposals, the outcome of such decision processes should not be further exploited to the advantage of successful applicants and the disadvantage of those not granted. Using citation analysis as one of the tools for the allocation of research budgets may be acceptable, but considering the amount of grants obtained by individual scientists as a parameter of esteem is unjustified and should be discouraged. Scientists with large grants are already rewarded in the sense that they have more financial possibilities than scientists with equal capacities of which research proposals are not granted. The mere fact that research proposals are granted should not be rewarded a 'second time'. On the contrary, obtained financial support should appear in the denominator, not the nominator, of any parameter of research quality. Unfortunately, the latter, not the former, is becoming practice in some Dutch University Medical Centres, such as the Academic Medical Center in Amsterdam.<sup>11</sup> It is remarkable that this very simple consideration is not broadly shared. It would be hard to find a company which considers the sum of investments, salaries and income from its sold products as the total profit of the company. We are not stating that those with ten times as much money (grants) should have a ten times higher output. We are only saying that the issue should not be ignored. There is a tendency in academic hospitals to select leading scientists (principal investigators) on the basis of success in obtaining such grants rather than on the success of their work. Such a policy seems primarily driven by the aim of academic hospitals to cut down their own expenses on research and to move these investments to third parties (see also the section 'Citation analysis'). This is not good news for fundamental research.

### Conclusion

Quality assessment of scientific output is a heavily disputed issue. Here we describe the application of a novel method with an obvious advantage over existing ones. The *h-index* is more simple to calculate and is certainly more relevant than using the impact factor of scientific journals for individual papers and authors and may become a suitable parameter for 'past performance' of an individual scientist. The variability in *h-index* between individual professors in clinical cardiology is substantial. At this stage we would hesitate to interpret the observed differences as differences in scientific quality only. For the application of this simple method to the scientific output of other groups of scientists, it is important that variability between different fields is investigated. For the development of an acceptable system of quality assessment it is vital that the applied method is acceptable for the members

of the scientific community, not for 'managers of science'.

### Limitation

We hesitated about the question whether or not we should publish the data in table 2 anonymously. We wish to emphasise that the large differences that we observed between individuals cannot be interpreted as quality differences between those individuals without further specialised research, which should focus on differences between (sub)fields. It is very well possible that differences between subfields, even within clinical cardiology, underlie the observed variance to a large degree and that the observations do not point to differences in quality *per se*. This has thus far not received sufficient attention and without more in-depth analyses, exercises such as the one we have performed, may be unfair and may potentially harm the careers of those labelled as 'less strong'.

We would certainly not publish such an analysis of the scientific output of younger scientists at the beginning of their careers without more research. In this case we thought that the information is of sufficient interest and the fact that the individuals have a completed career in the sense that they have reached the top of the academic ranks, played a role as well. It goes without saying that the large interindividual differences as observed for professors in clinical cardiology, should be an important warning for those who wish to compare scientists of different disciplines. To mention one pitfall, the ongoing research of one of us (TO) clearly shows that the citation frequency of clinical cardiological oriented papers is higher than that of basic cardiological oriented papers. Ignoring such pivotal information may create a future for academic hospitals with a staff with a majority of overestimated clinical researchers and a minority of underestimated basic scientists. ■

### References

- 1 Van der Heyden MAG, Derks van de Ven T, Opthof T. Fraud and misconduct in science: the stem cell seduction. *Eur Heart J* 2009;17:25-9.
- 2 Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005;294:218-28.
- 3 Cole S. Citations and the evaluation of individual scientists. *Trends Biochem Sci* 1989;14:9-13.
- 4 Opthof T. Sense and nonsense about the impact factor. *Cardiovasc Res* 1997;33:1-7.
- 5 MacRoberts MH, MacRoberts BR. Citation analysis and the science policy arena. *Trends Biochem Sci* 1989;14:8-12.
- 6 Opthof T, Coronel R. The most frequently cited papers of *Cardiovasc Res* (1967-1998). *Cardiovasc Res* 2000;45:3-5.
- 7 Seglen PO. From bad to worse: evaluation by Journal Impact. *Trends Biochem Sci* 1989;14:326-7.
- 8 Adler R, Ewing J, Taylor P. Joint Committee on Quantitative Assessment of Research. Citation Statistics, 2008. <http://www.mathunion.org/Publications/Report/CitationStatistics>.
- 9 Leydesdorff L. Caveats for the use of citation indicators in research and journal evaluations. *JASIS* 2008;59:278-87.
- 10 Opthof T, Coronel R, Piper HM. Impact factors: no totum pro parte by skewness of citation. *Cardiovasc Res* 2004;61:201-3.
- 11 Anonymus. Advice of the Research Council on the evaluation of the AMC research 2008. Internal report, 2009.
- 12 Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* 2005;102:16569-72.
- 13 Bornmann L, Mutz R, Daniel HD. Are there better indices for evaluation purposes than the *h* index? A comparison of nine different variants of the *h* index using data from biomedicine. *JASIS* 2008;59:830-7.
- 14 Ball P. Index aims for fair ranking of scientists. *Nature* 2005;436:900.
- 15 Opthof T, Coronel R. The impact factor of leading cardiovascular journals: where is your paper best cited? *Neth Heart J* 2002;10:198-202.
- 16 Liang L. *h*-index sequence and *h*-index matrix: Constructions and applications. *Scientometrics* 2006;69:153-9.
- 17 Jin B. The AR-index: complementing the *h*-index. *ISSI Newsletter* 2007;3:6.
- 18 Van Raan AJF. Statistical properties of bibliometric indicators: research group indicator distributions and correlations. *JASIST* 2006;57:408-30.
- 19 Opthof T, Coronel R, Janse MJ. The significance of the peer review process against the background of bias: priority ratings of reviewers and editors and the prediction of citation, the role of geographical bias. *Cardiovasc Res* 2002;56:339-46.
- 20 Moed HF, Burger WJM, Frankfort JG, Van Raan AJF. A comparative study of bibliometric past performance analysis and peer judgement. *Scientometrics* 1985;8:149-59.
- 21 Moed HF. Citation analysis in research evaluation. Springer, Dordrecht, The Netherlands, 2005, pp: 229-57.
- 22 Van Raan AJF. Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics* 2006;67:491-502.
- 23 Van den Besselaar P, Leydesdorff L. Past performance of successful grant applications. The Hague. Rathenau Institute. SciSA Report 0704, 2007.
- 24 Cichetti DV. The reliability of peer review for manuscript and grant submissions: a cross-disciplinary investigation. *Behav Brain Sci* 1991;14:119-86.